PENAKSIRAN FUNGSI DENSITAS TIPE KERNEL DENGAN METODE *CROSS-VALIDATION* (C-V)

(*Kernel Type Density Function Estimates with Cross-Validation Methods*)

MOZART WINSTON TALAKUA
Jurusan Matematika FMIPA Universitas Pattimura Ambon
Jl. Ir. M. Putuhena, Kampus Unpatti, Poka-Ambon

ABSTRACT

The choice of the bandwidth *h* is the main problem of kernel density function. In some situations it might be quite useful to have a set of retes corresponding to different bandwidth.It is necessary to agree which bandwidth is an appropriate one. Cross-Validation (C-V) is a well-known method to optimize the smoothing parameter *h*. In this research, we will analysis about three methods of Cross-Validation: *Maximum Likelihood Cross-Validation, Least-Square Cross-Validation (Unbiased Cross-Validation) and Biased Cross-Validation* with bandwidth choice optimal.

Keywords : *optimal bandwidth, Maximum Likelihood Cross-Validation, Unbiased Cross-Validation, Biased Cross-Validation*

PENDAHULUAN

Pemilihan parameter penghalus *(Bandwidth)* *h* merupakan masalah pokok dari estimator densitas tipe kernel. Pada beberapa situasi, mungkin cukup berguna untuk memakai seperangkat estimasi yang menghubungkan dengan *bandwidth* yang berbeda. Estimasi-estimasi tersebut dapat menonjolkan aspek-aspek yang berbeda dalam struktur data, tetapi penyajian dan interpretasi dari kurva semacam itu cukup subjektif. Adalah perlu untuk mencapai kesepakatan mengenai *bandwidth* yang mana yang tepat, karena dua kondisi yang saling kontradiktif dalam pemilihan *h* yaitu sifat bias mengharuskan memilih *h* yang kecil untuk memperkecil bias sedangkan variansi mengharuskan memilih *h* yang besar untuk memperkecil variansi. Metode penaksiran yang memperbaiki kelemahan penaksiran dengan histogram adalah metode penaksiran kernel yang didefinisikan dengan k_h adalah fungsi pembobotan pada data pengamatan X_{i_i}, .i= 1,2,….,n dan parameter bandwidth *h* , dimana *bandwidth h* adalah suatu parameter penghalus dalam fungsi kernel yang berfungsi untuk mengatur tingkat kehalusan fungsi kernel tersebut. Metode penaksiran ini mudah dan sederhana dalam penaksirannya, tetapi sifat-sifat analitiknya belum dibuktikan secara lengkap, maka perlu dalam penelitian ini akan dibuktikan sifat-sifat penaksir kernel secara lengkap. Penaksir kernel ini didefinisikan sebagai $\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right)$, dimana $X_1, X_2, \ldots, X_n$ adalah sample pengamatan dan *k* adalah fungsi kernel dengan sifat : $\int_{-\infty}^{\infty} k(x)d(x) = 1$, serta *h* adalah parameter *bandwidth*

Ide estimasi kernel dikenalkan oleh Murray Rosenblatt (1956) yang memperbaiki histogram. Sejauh itu ada beberapa tulisan yang mengembangkan estimasi densitas tipe kernel dengan sifat-sifat asimtotik seperti Rosenblatt pada tahun 1971. Seheult & Quesabberry (1971) mengatakan bahwa estimasi kernel adalah bias, begitupun dengan Parsen (1962) dan Bartlett (1963), sedangkan Singho Shirahatta dan In-Sun Chu (1991) menerangkan sifat *Integral Square Error* (*ISE*) dan estimasi kernel. W.H. Swanepoel (1988) menemukan sifat-sifat *Asimtotik Mean Integral Square Error (A-MISE)* dari suatu estimasi kernel *f* bila *f* diasumsikan kontinu dimana-mana. Atau kontinu kecuali pada sejauh berhingga titik-titik diskontinu. Talakua (2001) mengembangkan Penaksiran Fungsi densitas tipe Kernel untuk daerah (0,~). Metode–metode tersebut hanya dioperasikan pada *f* yang diketahui dan untuk *f* yang tidak deketahui metode tersebut di atas tidak cukup untuk menyelesaikan permasalahannya, sehingga ide dan gagasan baru muncul dalam menyelesaikan persoalan estimasi densitas tipe kernel dengan Metode *Cross-Validation (C-V)*. Dalam Penelitian ini akan dibahas tiga metode *Cross-Validation* yaitu : *Maximum Likelihood Cross-Validation, Least-Square Cross-Validation (Unbiased Cross-Validation) dan Biased Cross-Validation* dengan pemilihan *bandwith* yang optimal.

Penelitian ini diharapkan dapat memberikan manfaat antara lain, dari segi subjektif peneliti, kegunaan yang dapat diambil adalah untuk memperkuat konsep dan pemahaman peneliti tentang konsep analisis Statistika matematika yaitu penaksiran densitas tipe kernel dengan metode *Cross-Validation*, dari segi pengembangan Pendidikan Tinggi di Indonesia, khususnya Jurusan Matematika FMIPA UNPATTI diharapkan dapat memberikan sumbangan pemikiran untuk semakin memperluas dan mendalami konsep dari metode *Cross-Validation* secara praktis maupun teoritis.

Akhirnya Dengan adanya penelitian ini, maka nantinya dapat diadakan penelitian lebih lanjut dengan melibatkan berbagai metode untuk mendapatkan hasil yang beraneka ragam dan semakin mendalami mengenai penaksiran densitas tipe kernel.

METODE PENELITIAN

Penelitian yang dikerjakan ini bersifat studi literatur sehingga dalam prosesnya diperlukan dan dilakukan beberapa langkah sebagai berikut:

a. Bahan atau Materi Penelitian

Bahan dan materi yang digunakan dalam penelitian ini berupa karya ilmiah para matematikawan dan statistikawan, baik yang disajikan dalam seminar maupun yang dimuat dalam jurnal, buletin, buku cetak (*textbook*) dan informasi yang diperoleh melalui internet.

b. Alat Penelitian

Alat yang digunakan untuk penelitian ini adalah seperangkat komputer yang mendukung proses penelitian ini.

c. Cara Penelitian

Penelitian ini dilakukan dengan cara mengumpulkan, mempelajari dan menganalisis karya-karya ilmiah (jurnal dan buletin) atau buku-buku (*textbook*) yang berkaitan dengan permasalahan yang diteliti, kemudian diuraikan dan disajikan secara benar (*mathematically correct*) dalam bentuk tulisan yang logis dan runtut.

d. Analisis Hasil Penelitian

Analisis dan pengujian hasil penelitian yangv telah dibuat dalam bentuk tulisan yang logis dan runut dengan cara mengkomunikasikan dan mendiskusikan dengan metematikawan dan statistikawan lain yang memiliki ilmu yang sama, selanjutnya dibuat dalam bentuk laporan penelitian.

HASIL DAN PEMBAHASAN

Kernel k dengan bandwidth h dinotasikan sebagai : $k_h(x) = \frac{1}{h} k\left(\frac{x}{h}\right)$ dengan parameter bandwidth *h*. Fungsi kernel adalah fungsi pembobotan pada data pengamatan $(X_i), i = 1,2,\ldots,n$ dengan parameter bandwidth *h* Parameter bandwidth *h* adalah suatu parameter penghalus, yang dalam penaksiran kernel ini berfungsi untuk mengatur tingkat kehalusan fungsi kernel tersebut, juga dapat dikatakan sebagai jangkauan titik-titik pengamatan *X* yang akan dipengaruhi oleh data.Pembobotan dilakukan setelah peubah acak yang akan diamati diurutkan terlebih dahulu dari yang kecil sampai yang terbesar.

Fungsi kernel ini memenuhi beberapa kriteria, yaitu:

A_1. Fungsi kernel simetrik di titik nol, dan integralnya sama dengan satu

$$k(-x) = k(x)$$

A_2. Fungsi kernel $\int k(x) d(x) = 1$, maka kernel merupakan fungsi kepadatan.

Dilihat dari kedua sifat tersebut di atas maka fungsi kepadatan peluang (fkp) dapat ditaksir dengan kernel. Selanjutnya penaksir kernel $\hat{f}_h$ untuk f didefinisikan sebagai rataan nilai kernelnya dan ditulis sebagai berikut :

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - X_i) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_1}{h}\right)$$

(3.1)

dimana $X_i, i = 1,2,\ldots,n$ adalah peubah acak yang diamati.

Maximum Likelihood Cross-Validation.

Pandang estimasi-estimasi densitas kernel $\hat{f}_h(x)$ dan misalkan ingin diuji untuk h yang spesifik dari hipotesis

$$\hat{f}_h(x) = f(x) \text{ Vs } \hat{f}_h(x) \neq f(x)$$

Likelihood Rasio Test akan didasarkan pada test statistik $\frac{f(x)}{f_h(x)}$. Bagi suatu bandwidth yang baik, statistik ini seharusnya dekat ke 1 Dapat juga dikatakan bahwa pada rata-rata (atas X) $E_x\left[\log\left(f/f_h\right)(x)\right]$ seharusnya 0. dengan begitu suatu bandwidth yang bagus, yang meminimalkan ukuran keakurasian, pada efeknya mengoptimalkan informasi Kulback-Leibler:

$$d_{kl}\left(f, \hat{f}_h\right) = \int \log\left(\frac{f}{f_h}\right)(x) f(x) dx \qquad (3.1.1)$$

dengan mengingat bahwa $d_{kl}\left(f, \hat{f}_h\right)$ adalah jarak, bukan metrik, sebab pada umumnya $d_{kl}\left(f, \hat{f}_h\right) \neq d_{kl}\left(\hat{f}_h, f\right)$

Didasarkan pada ‘’ *leave-one-out estimate*’’

$$\hat{f}_{h,i}(x) = (n-1)h^{-1}\sum_{j\neq i} k\left(\frac{x_i - x_j}{h}\right) \qquad (3.1.2)$$

Kemudian didefinisikan

$$\prod_{i=1}^{n} \hat{f}_{h,i}(x_i) = (n-1)^{-n} h^{-n} \prod_{i=1}^{n}\sum_{j\neq i} k\left(\frac{x_1 - x_j}{h}\right) \qquad (3.1.3)$$

Apabila Persamaan (3.1.3) dilogaritmakan dan dinormalisasikan dengan faktor n^{-1} maka didefinisikan *Maximum Likelihood Cross Validation (ML-CV)* sebagai berikut :

$$CV_{KL}(h) = n^{-1}\sum_{i=1}^{n} Log\left[\sum_{j\neq i} k\left(\frac{x_i - x_j}{h}\right)\right] - \log[(n-1)h] \qquad (3.1.4)$$

Dengan demikian $\hat{h}_{KL}$ dikatakan *“baik”* apabila mendekati maksimum yang finit dari $CV_{KL}(h)$, yaitu :

$$\hat{h}_{KL} = \arg\max_h CV_{KL}(h)$$

,selanjutnya dihubungkan pada informasi ***Kullback-Leiber*** dan diasumsikan bahwa X_i berdistribusi Identik.

Least-Square Cross-Validation.

Pandang suatu ukuran jarak alternatrif antara $\hat{f}$ *dan* f yaitu " *Integrated Square Error"* (*ISE)* yang didefinisikan :

$$ISE(h) = \int \left(\hat{f_h} - f \right)^2 (x) dx \qquad (3.2.1)$$

$$= \int \left[\hat{f_h}(x) \right]^2 dx - 2\int \left(\hat{f_h} f \right)(x) dx + \int f^2(x) dx$$

(3.2.2)

Apabila Persamaan (3.2.2) dikurangi dengan bagian konstan, mengakibatkan peminimalan *ISE* dalam hubungannya dengan h yang ekuivalen dengan meminimalkan :

$$ISE(h) - \int f^2(x) dx =$$

$$= \int \left[\hat{f_h}(x) \right]^2 dx - 2\int \left(\hat{f_h} f \right)(x) dx$$

(3.2.3)

Untuk mengestimasi pada bagian ini digunakan "*Leave –One-Out estimation*"

$$E_x \left[\hat{f_h}(x) \right] = n^{-1} \sum_{i=1}^{n} \hat{f}_{h,i}(x_i) \qquad (3.2.4)$$

Dengan demikian apabila digunakan estimasi ini berarti untuk menentukan suatu *Bandwidth* yang baik yang meminimalkan ruas kanan dari persamaan (3.2.3) dengan menggunakan estimasi pada persamaan (3.2.4) maka akan didefinisikan *Least Square Cross Validation* sebagai berikut :

$$CV(h) = \int \hat{f_h}^2 (x) dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{h,i}(x_i) \qquad (3.2.5)$$

Berdasarkan persamaan (3.1.5.) maka *bandwidth* tertentu pada fungsi kernel secara eksak diperoleh :

$$E(CV(h)) = E[ISE(h)] + 2\left[E_x \left(\hat{f_h}(x) \right) \right] - E\left[\frac{1}{n} \sum \hat{f}_{h,i}(x) \right] - \|f\|_2^2$$

$$= MISE\left[\hat{f_h} \right] - \|f\|_2^2 \qquad (3.2.6)$$

Dari persamaan (3.1.6) Inilah *Least Square Cross-Validation* disebut **Unbiased Cross-Validation.** Jadi untuk pembahasan selanjutnya persamaan (3.1.5) ditulis :

$$UCV(h) = \int \hat{f_h}^2 (x) dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{h,i}(x_i) \qquad (3.2.7)$$

Suatu bandwidth menjadi optimal secara asimtotik, jika

$$\frac{ISE\left(\hat{h}_n \right)}{Inf_{h>0} ISE(h)} \xrightarrow{a.s} 1 \quad (n \longrightarrow \infty) \qquad (3.2.8)$$

Teorema 3.2

Jika

$$\sup_{hh'} \left| \frac{(ISE(h) - ISE(h') - UCV(h) - UCV(h')}{ISE(h) + ISE(h')} \right| \xrightarrow{a.s} 1$$

$(n \longrightarrow \infty)$ Maka h_{UCV} optimal secara asimtotik

Bukti:

Dari persamaan 3.2.1 diperoleh

$\hat{h}_{ISE} = \arg_h \min ISE(h)$ yang berarti

$0 < ISE(\hat{h}_{UCV}) - ISE(\hat{h}_{ISE})$, sedangkan persamaan 3.2.5. diperoleh $\hat{h}_{UCV} = \arg_h \min UCV(h)$ yang berarti $0 < UCV(\hat{h}_{UCV}) - UCV(\hat{h}_{ISE})$ sehingga untuk $(n \longrightarrow \infty)$, $\varepsilon > 0$ berlaku

$$0 < \frac{ISE(\hat{h}_{UCV}) - ISE(\hat{h}_{ISE}) - [UCV(\hat{h}_{UCV}) - UCV(\hat{h}_{ISE})]}{ISE(\hat{h}_{UCV}) + ISE(\hat{h}_{ISE})} \le \varepsilon$$

dengan demikian diperoleh

$$0 \ge UCV(\hat{h}_{UCV}) - UCV(\hat{h}_{ISE}) \ge (1-\varepsilon) ISE(\hat{h}_{UCV}) - (1+\varepsilon) ISE(\hat{h}_{ISE})$$

dengan menghilangkan unsur

$UCV(\hat{h}_{UCV}) - UCV(\hat{h}_{ISE})$ sehingga didapatkan

$\dfrac{(1+\varepsilon)}{(1-\varepsilon)} \ge \dfrac{ISE(\hat{h}_{UCV})}{ISE(\hat{h}_{ISE})}$ karena berlaku untuk setiap ε

dengan demikian $\lim\limits_{n \to \infty} \dfrac{ISE(\hat{h}_{UCV})}{ISE(\hat{h}_{ISE})} = 1$ yang berarti bahwa h_{UCV} optimal secara asimtotik

Biased Cross Validation

Teknik ini pertama kali ditemukan oleh Scott Terrell (1987). Gagasan dari teknik pemilihan *bandwidth* ini didasarkan pada estimasi langsung dari *A-MISE* $\left(\hat{f_h} \right)$ densitas kernel yang disajikan sebagai berikut :

$$A\text{-}MISE \left(\hat{f_h} \right) = \frac{1}{nh} \|k\|_2^2 + \frac{h}{4} \mu_2^2(k) \|f''\|_2^2$$

(3.3.1)

Sehingga Biased Cross Validation didefinisikan sebagai :

$$BCV_1(h) = \frac{1}{nh} \|k\|_2^2 + \frac{h}{4} \mu_2^2(k) \|f''\|_2^2$$

(3.3.2)

dari metode ini ditemukan bahwa untuk mengecilkan $MISE\left[\hat{f_h} \right]$ seharusnya memilih urutan bandwidth yang proporsional terhadap $n^{-1/5}$. Dengan lemma berikut:

Lemma 3.3

Diberikan $\hat{f}_h(x)$ pada persamaan 3.1 maka diperoleh

$$Var\left[\hat{f}_h''(x)\right] = h^{-5}n^{-1}\left\|k''\right\|_2^2 + o(h^{-5}n^{-1}) \sim n^{-1}h^{-5}\left\|k''\right\|_2^2$$

$$A\text{-}MISE\left(\hat{f}_h''(x)\right) =$$

$$h^{-5}n^{-1}\left\|k''\right\|_2^2 \quad + \quad \frac{h}{4}(f^{(iv)}(x)\mu_2^2(k) \qquad (3.3.4)$$

Dengan demikian dari (3.3.4) terlihat bahwa variansi dari $\left(\hat{f}_h''(x)\right)$ tidak mendekati 0 untuk pemilihan $h \sim n^{-1/5}$.

Selanjutnya Scott & Terell (1987) memperlihatkan bahwa $\hat{h}_{BCV}$ adalah optimal secara asimtotik sebagai yang disajikan melalui teorema berikut:

Teorema 3.3
Jika

$$\underset{hh'}{Sup}\left|\frac{(ISE(h) - ISE(h') - [BCV(h) - BCV(h')}{ISE(h) + ISE(h')}\right| \xrightarrow{a.s} 1$$

$(n \longrightarrow \infty)$ Maka h_{BCV} optimal secara asimtotik

Bukti:

Dari persamaan 3.2.1 diperoleh $\hat{h}_{ISE} = \arg_h \min ISE(h)$ yang berarti $0 < ISE(\hat{h}_{BCV}) - ISE(\hat{h}_{ISE})$, sedangkan persamaan 3.2.5. diperoleh $\hat{h}_{BCV} = \arg_h \min BCV(h)$ yang berarti $0 < BCV(\hat{h}_{BCV}) - BCV(\hat{h}_{ISE})$ sehingga untuk $(n \longrightarrow \infty)$, $\varepsilon > 0$ berlaku

$$0 < \frac{ISE(\hat{h}_{BCV}) - ISE(\hat{h}_{ISE}) - [BCV(\hat{h}_{BCV}) - BCV(\hat{h}_{ISE})]}{ISE(\hat{h}_{BCV}) + ISE(\hat{h}_{ISE})} \leq \varepsilon$$

dengan demikian diperoleh

$$0 \geq BCV(\hat{h}_{BCV}) - BCV(\hat{h}_{ISE}) \geq (1-\varepsilon)ISE(\hat{h}_{BCV}) - (1+\varepsilon)ISE(\hat{h}_{ISE})$$

dengan menghilangkan unsur $BCV(\hat{h}_{BCV}) - BCV(\hat{h}_{ISE})$ sehingga didapatkan

$$\frac{(1+\varepsilon)}{(1-\varepsilon)} \geq \frac{ISE(\hat{h}_{BCV})}{ISE(\hat{h}_{ISE})}$$ karena berlaku untuk setiap ε

dengan demikian $\lim\limits_{n\to\infty} \dfrac{ISE(\hat{h}_{BCV})}{ISE(\hat{h}_{ISE})} = 1$ yang berarti bahwa h_{BCV} optimal secara asimtotik

Selanjutnya dari ketiga metode *Cross-Validation* tersebut, dapat dilihat tingkat *Asimtotik* sehingga untuk mengestimasi Fungsi densitas Tipe Kernel dapat dikembangkan untuk metode *Cross-Validation* yang cocok untuk suatu distribusi tertentu.
Sebagai contoh dapat di lihat pada pengambilan *n* sampel yang berbeda pada distribusi Binomial, sebagai berikut:

n	$E(x$	σ	h_{opt}		h_{Sbrg}	$A-MISE$		
			Uniform	**Triangle**		**Uniform**	**Triangle**	**Sbrg**
50	3.44	1.280	0.807	0.889	0.5	0.037	0.97	0.114
100	3.76	1.231	0.616	0.678	0.5	0.022	0.57	0.104
150	3.62	1.173	0.513	0.564	0.5	0.016	0.39	0.101
200	3.66	1.282	0.469	0.517	0.5	0.014	0.33	0.099
250	3.64	1.224	0.451	0.183	0.5	0.012	0.097	0.098
300	3.56	1.201	0.417	0.159	0.5	0.010	0.085	0.097

Dari tabel di atas terlihat bahwa jika pengambilan h_{Sbrg}, maka nilai $A-MISE$ seperti pada tabel di atas**.** Dapat dilihat perbandingannya dengan pengambilan h_{opt} penaksir kernel maka nilai $A-MISE_{h_{Unif}} < A-MISE_{h_{tri}} < A-MISE_{h_{Sbrg}}$ sehingga dapat disimpulkan bahwa pengambilan h_{opt} penaksir kernel Uniform lebih asimtotik dari penaksir kernel triangle dan h_{opt} sebarang**.**

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan penjelasan pada bagian sebelumnya maka dapat disimpulkan bahwa:

1. Pemilihan *bandwith* yang optimal untuk metode *Maximum Likelihood Cross- Validation(ML-CV)* yaitu $\hat{h}_{KL}$ dikatakan *"baik"* apabila mendekati maksimum yang finit dari $CV_{KL}(h)$, yaitu :
   $$\hat{h}_{KL} = \arg \max_h CV_{KL}(h)$$
2. Pemilihan *bandwith* yang optimal untuk metode *Least-Square Cross- Validation (Unbiased Cross-Validation)* yaitu jika
   $$\underset{hh'}{Sup}\left|\frac{(ISE(h) - ISE(h') - UCV(h) - UCV(h')}{ISE(h) + ISE(h')}\right| \xrightarrow{a.s} 1 \; (n \longrightarrow \infty)$$
   Maka $\hat{h}_{UCV}$ optimal secara asimtotik
3. Pemilihan *bandwith* yang optimal untuk metode *Bias Cross-Validation* yaitu Jika
   $$\underset{hh'}{Sup}\left|\frac{(ISE(h) - ISE(h') - [BCV(h) - BCV(h')}{ISE(h) + ISE(h')}\right| \xrightarrow{a.s} 1$$
   $(n \longrightarrow \infty)$ Maka $\hat{h}_{BCV}$ optimal secara asimtotik

4. Pengambilan h_{opt} penaksir kernel Uniform lebih asimtotik dari penaksir kernel triangle dan h_{opt} sebarang**,** karena nilai

$$A-MISE_{h_{Unif}} < A-MISE_{h_{tri}} < A-MISE_{h_{Sbrg}}$$

Saran

Diharapkan dapat dilakukan penelitian lanjutan untuk penaksiran fungsi densitas tipe kernel dengan metode yang lain, dengan jenis kernel yang berbeda

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Proyek Peningkatan Penelitian Pendidikan Tinggi,Direktorat Pendidikan Tinggi,Departemen Pendidikan Nasional, sesuai dengan surat Perjanjian Pelaksanaan Penelitian Dosen Muda,Studi kajian wanita dan Sosial Keagamaan Nomor: 127/P4/DPPM/DM,SKW,SOSAG/III/2004 tanggal 25 Maret 2004

DAFTAR PUSTAKA

Apostol, T.M.(1967), *Calculus Volume I,* Wiley International Edition.

Devroye and Penrod,C.S, (1984), *The Consistency of Kernel density Estimates*, Journal The Annalisis Statistics,Vol 12.,No..4, 1231-1249

Everitt,Brian S. (1994), *A Handbook of Statistical analyses using S-Plus*, New York.Chapman&Hall

Gaser.T,Muller.H.G. Mamitcsch Society, *Kernel for Nonparametric Curve Estimation*, Journal of The Royal Statistic Society, Seri B, 47 236-252

Hardle,W (1990), *Smoothing Tecnique With Implementation in S.*, Spinger Verlag, New York.

Hill, P.D. (1985), *Kernel Estimation of Distribution Finction*, Comm, Statist.Theory Methods, 14, 605-620

Jhon Rice (1984), *Bandwidth Choice For Non Parametric Regression*, Journal The Anall. Of Statistics, Volume 12, N0.4, 1215-1230

Jeffrey D. Hart and Thomas E. Wehrly, (1992), *Kernel regression When The Boundary region in Large, With an Application to testing the Adequacy of Polinomial Medels*, Journal of The American Statistical Assosiation, Vol.87.No.420. 1018-1024

Lethold,L (1976), *The Calculus With Analitic Geometry*, New York Harper International Edition.